Universitat Pompeu Fabra Barcelona

Treball de fi de màster Acadèmic

Neural Machine Translation of Discontinuous Units: A Case Study of Phrasal Verbs

Nom i Cognoms Nataliia Zolotukhina

Màster: Lingüística Teòrica i Aplicada

Edició: 2020-2021

Directors: Dra. Louise McNally

Any de defensa: 2021

Col·lecció: Treballs de fi de màster

Departament de Traducció i Ciències del Llenguatge

ABSTRACT

Neural machine translation models are mostly evaluated by automatic metrics such as BLEU. These metrics mainly catch the overlaps of words and phrases in machine translation and professional human translation of the same text, but they do not tell us whether the output text is grammatically correct or semantically identical to the input text. So how well do the neural models actually translate? We set out to answer this question for the language pair German-English, and we challenge the model with a specific pattern – phrasal verbs. They are interesting for us because they form one lexical item that consists of two discontinuous units. Does the model recognize that they belong together?

We conduct our analysis on 119 sentences translated by the Facebook FAIR model by Ng et al. (2019). Our findings suggest that the model is more likely to generate phrasal verbs in English when it receives German verbs with separable prefixes as an input. The qualitative analysis shows that the model produces grammatically correct sentences that correspond semantically to the source sentences. Moreover, it demonstrates an ability to learn certain syntactic patterns, capture semantics, and adapt to the context.

KEY WORDS: neural machine translation, phrasal verbs (English), separable verbs (German), Transformer, tokenization, BLEU metric, discontinuous morphology.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to Dr. Louise McNally for guiding me throughout this project, sharing her expertise, resolving my doubts and overcoming the obstacles together with me. The completion of this work would not have been possible without her continuous support and valuable recommendations.

I would like to extend my sincere thanks to Dr. Denis Paperno for the original idea of this project. Although it has undergone many changes in the process, he set up the direction of this research.

I am deeply indebted to Mariana Romanyshyn for her extensive knowledge in the field of computational linguistics, insightful suggestions over my topic, genuine support and openness to help with anything.

I am also grateful to my good friend Daniel Schmitz for consulting me over the German language during all these months, and to my classmates Daniela Corbetta and Julia Pardo for all our discussions, sharings, practical suggestions, and mutual encouragement in times of struggles.

Finally, I am eternally thankful to my family for their profound belief in my abilities and for cheering me up when I doubted myself.

TABLE OF CONTENTS

1. INTRODUCTION

Machine translation has undergone a long process of development, from rule-based approaches, through statistical machine translation solutions, and up until deep learning methods. The quality of translation becomes better, but do the machines really begin to understand human language? In this study we aim to look into the quality of machine translation and to see whether our trust in it is justified.

We choose one of the newest neural machine translation (NMT) models – the model from Facebook by Ng et al. (2019), for the German-English language pair. We know about this model that it was the winner of the WMT19 shared news translation task at the Fourth Conference on Machine Translation, and that it scores 40.8 on BLEU. But what is behind this score? The general question we pose in this study is how well the model translates from German to English. Since this is a very general question, we choose a certain language pattern to focus on in our evaluation – phrasal verbs.

The motivation behind this choice comes from the background processes of how the models generally operate. They process text as a sequence of continuous characters that correspond to words or subwords. But what happens when a sequence that forms a single lexical item is interrupted in text? Does a model recognize it as a single item anyway? Phrasal verbs are a suitable pattern to challenge the model with interrupted sequences, and in this way, we pose a more specific research question: does an NMT model connect the discontinuous units that belong together, or, in our specific case, how does it handle phrasal verbs? While answering this question, we will also pursue our aim of figuring out how well the model translates in general.

2. BACKGROUND

Natural language processing, as it is nowadays, relies largely on deep learning algorithms. Particularly in the field of machine translation, traditional statistical machine translation (SMT) models (e.g., Koehn et al., 2003) have been replaced by better performing neural machine translation (NMT) models (e.g., Sutskever et al., 2014). In this section, we will review the characteristics of the NMT models, the state-of-the-art approaches, essential related concepts such as tokenization and byte-pair encoding, and the most relevant related studies in the field.

2.1 Neural Machine Translation

A standard algorithm for the NMT models is based on the encoder-decoder architecture (Jurafsky and Martin, 2020, p.203). The encoder is meant to take a source sentence and encode it into a vector representation; a decoder then uses this representation to generate a target sentence, i.e., a translation (Cho et al., 2014). An important component in this process is an embedding layer that maps words into word embeddings (vector representations) – mathematical objects that machines can read and operate on (Goldberg, 2017, p.3).

As summarized by Jurafsky and Martin (2020, p.203), there are two ways to implement the encoder-decoder network: either with recurrent neural networks (RNNs; Elman, 1990) or with Transformers (Vaswani et al., 2017). Initially, the dominant NMT models were based mainly on RNNs, particularly on the long short-term memory approach (LSTM; Hochreiter and Schmidhuber, 1997). RNNs are suitable to work with sequential data because they take a sequence of items and produce its "informative representation", which the other components of the model can operate on (Goldberg, 2017, pp.3-4). Both the encoder and the decoder components are separate RNNs, and as discussed in Chung et al. (2014, p.2), this approach proved to work well on sequence-to-sequence tasks like machine translation and to capture long-term dependencies to some extent. However, Cho et al. (2014) revealed that the RNN-based models for machine translation work well only on short sentences with known vocabulary; with longer sentences or with more unknown words, the performance of the models decreases.

This problem was addressed by Bahdanau et al. (2014), who suggested an extension component to the encoder-decoder scheme – the attention mechanism. The idea behind the attention mechanism was to focus only on relevant information: for each output word, a model searches the source sentence for the positions associated with it; to produce its translation, the model takes into account these relevant source positions and previously generated target words (Bahdanau et al., 2014). With such a workflow, the decoder receives more information and is able to adapt to what is relevant at each step. This leads to an increase in the performance of NMT models on longer sentences (Bahdanau et al., 2014).

In 2017, a novel NMT architecture was introduced by Vaswani et al. (2017) – the Transformer. The novelty of the Transformer is that it uses an encoder and a decoder, but it excludes RNNs from them and relies solely on attention. In the original Transformer, the encoder and the decoder are both composed of 6 identical layers, and each layer contains one (on the encoder side) or two (on the decoder side) attention mechanisms. The crucial point of the Transformer's attention is that it is not a single function; it has multiple parallel layers of attention (8 in the original paper) and therefore is called multi-head attention. As described by the authors, the advantage of this approach is that at every position, the attention has access to different subspaces and can attend to all relevant positions of the encoder and decoder simultaneously (Vaswani et al., 2017, pp.4-5). This brings major improvements compared to the previous solutions – the computation processes are now parallelized; thus, the amount of computation is reduced, so the models can be trained faster and perform quicker (Vaswani et al., 2017, pp.6-10). Moreover, the Transformer addresses the issue of long-term dependencies. Previous RNNs had to go back and forth through the whole sequence to learn long-term dependencies; with the parallelized Transformer algorithm that allows the processing of the whole sequence at once, the paths between the dependencies are decreased; hence, less information is lost in the process (Vaswani et al., 2017, p.6).

Being able to address the limitations and the pitfalls of the previous solutions, the Transformer has become a new state-of-the-art approach in many NLP tasks. Originally suggested for machine translation, it has also found many applications in language modeling, question answering, text summarization, etc. in models such as BERT by Google (Devlin et al., 2018), RoBERTa by Facebook (Liu et al., 2019), GPT by OpenAI (Radford et al., 2018), and others. A large collection of the pre-trained transformer-based models is available in the open-source Transformers library maintained by Hugging Face (Wolf et al., 2019).

2.2 Tokenization

Besides the architecture and the algorithm of an NMT model, it is important in what format the language data are given. Machines cannot read a running text and identify word boundaries; the text has to be segmented by means of tokenization (Jurafsky and Martin, 2020, p.16).

Depending on a language system, tokens can be represented as words (e.g., in English) or as characters (e.g., in Chinese). When we talk about word tokens, the most straightforward approach is to split text by white spaces or punctuation symbols. However, there are various things to be considered: internal punctuation in some words, dates, prices, email addresses, URLs (e.g., *Ph.D.*, *$10.00, 10.10.2020*), multiword units that should not be broken apart (e.g., New York), as contrasted to the contractions that should be separated (e.g., *don't, can't, I'm*), and others (Jurafsky and Martin, 2020, pp.16-18).

Different tokenizers address these questions differently, and it turns out that this influences the performance of an NMT model. Domingo et al. (2018) conducted a series of experiments on five different tokenizers over ten language pairs in order to evaluate the effect of tokenization on the quality of NMT. The inspected tokenizers were SentencePiece, Mecab, Stanford Word Segmenter, the OpenNMT tokenizer, and the Moses tokenizer. The authors reported a significant correlation between the tokenization and the overall quality of translation and demonstrated that the choice of the best tokenizer would depend on the language pair in use due to variations in language systems. In the present research, we work with German, and according to Domingo et al. (2018), the best tokenizer for German is Moses (p.8). Moses is a rule-based tokenizer included with the Moses toolkit (Koehn et al., 2007) – an open-source toolkit for statistical machine translation. It normalizes characters, separates word tokens from punctuation symbols, but leaves intact special tokens like URLs or dates (Domingo et al., 2018, p.3).

Segmenting text into words is essential not only for an NMT model to be able to read the text but also for learning the vocabulary. But one important problem in translation is that a model cannot learn all possible vocabulary; the model is trained on certain corpora, and they do not contain all existing words of a given language (Jurafsky and Martin, 2020, p.18). Sennrich et al. (2015) summarize that on average, NMT models have a limited vocabulary of 30.000-50.000 words, and they are not able to translate the words that are out of their vocabulary – a problem especially relevant for morphologically rich languages (p.1). Sennrich et al. (2015) state that in order to address the problem of rare words, NMT models "require mechanisms that go below the word level" (p.1) – for example, to the subword level.

The idea of the subword tokenization is to allow a model to learn by itself which word parts (subwords) are relevant and meaningful to compose its vocabulary; subwords can, but need not, correspond to morphemes like the suffixes *-er*, *-est* (Jurafsky and Martin, 2020, p.18). There are several widely used approaches to subword tokenization (Schuster and Nakajima, 2012; Kudo, 2018), but the simplest one is based on the Byte-Pair Encoding (BPE; Gage, 1994) algorithm and is proposed by Sennrich et al. (2015). The BPE tokenization starts from representing the whole vocabulary as separate characters, and in the process of iteration, it searches for the characters that co-occur most frequently in the text and combines them in the subwords (Jurafsky and Martin, 2020, pp.18-20). This approach copes effectively not only with unknown words, but also with known rare words in a model vocabulary, and thus it improves the overall performance of the model (Sennrich et al., 2015).

2.3 NMT Model Evaluation

Various combinations of model characteristics discussed above lead to different performance results and the translation quality of NMT models. The best evaluation of a machine translation model would be done by a human rater, but this may be a very expensive and time consuming process (Jurafsky and Martin, 2020, p.222). Instead, automatic metrics have been designed for these purposes. Among all available options, the most popular is the BLEU (BiLingual Evaluation Understudy) metric (Papineni et al., 2002).

The main idea of the BLEU metric is that a machine translation is good when it is close to a professional human translation; hence, it measures the overlaps of words and phrases of different length in the translations obtained from people and from machines (Papineni et al., 2002, p.1). The advantage of this metric is that it evaluates a model in any language quickly and at small cost (Papineni et al., 2002, p.1). But its disadvantage is that it does not fully reflect the translation quality (Wu et al., 2016, p.14). Callison-Burch et al. (2006) note that "under some circumstances an improvement in BLEU is not sufficient to reflect a genuine improvement in translation quality, and in other circumstances that it is not necessary to improve BLEU in order to achieve a noticeable improvement in translation quality" (p.249). Nevertheless, the BLEU metric is still the main reference point in neural machine translation, and the majority of researchers report the improvements in their models using the BLEU scores.

Thus, we have NMT models trained by machines and evaluated by machines, and the BLEU metric that is not fully informative. So do we just blindly trust the BLEU score? How do we know that a model actually translates well? A number of studies were carried out to explore whether neural NLP models hold any linguistic knowledge. Mareček and Rosa (2019) demonstrated that the Transformer captures syntax. The same was reported by Vig and Belinkov (2019) for the GPT-2 model; they found that different parts of speech in a sentence are targeted at different layers of the attention mechanism, and different attention heads capture specific patterns and distant dependencies. Jawahar et al.'s (2019) study on BERT showed that in addition to syntax, BERT also captures some semantic features.

Summarizing previous studies, we see that the state-of-the-art transformers hold some knowledge of language structure. They have a certain notion of syntax and capture distant dependencies in a sentence. On the other hand, they do not have a notion of lemma because the subword tokenization is applied – they read words as sequences of characters or byte pairs. But what happens when the sequence is interrupted, and a word is "broken"? In other words, what happens when lemma is not a single unit anymore? These questions underlie the motivation of the present study, and one way to address them is to appeal to discontinuous morphology.

2.4 Phrasal Verbs as Discontinuous Units

Morphemes are the smallest units in language that bear meaning and cannot be segmented further (Haspelmath, 2002, p.16). However, they can be interrupted by other units (Harris, 1945, p.121). If two or more continuous morphemes (that are not interrupted) occur together all the time, they form a new discontinuous morpheme (Harris, 1945, p.121). For this study, we will consider phrasal verbs as an example of discontinuous morphemes.

A phrasal verb in English is a combination of a verb and a particle "that functions as a single verb, both parts giving up meaning in order to form a new lexical item" (Darwin and

Gray, 1999, p.65). Some examples of phrasal verbs are *give up* (= *surrender*), *call off* (= *cancel*), *go on* (= *continue*), etc. So, both parts can exist on their own with different meanings, but when combined together, they become a new, single unit represented by two tokens. Moreover, a complement of the phrasal verb can be placed between the two tokens in a sentence and separate them from each other (Thim, 2012, pp.21-22): *They **called off** the meeting – They **called** the meeting **off***.

It is important to distinguish between phrasal verbs and verbs that select for prepositional phrases (Vestergard, 2019, p.3): *the roof **fell in** – he **fell** <u>in love</u>*. Verbs that select for prepositional phrases do not form a single lexical item, unlike phrasal verbs. The challenge here is that the same lexeme can be a preposition in one context and a phrasal verb particle in another context (Vestergard, 2019, p.4): *She switched **off** the light* (particle) – *The pen fell **off** the table* (preposition). One way to distinguish them is to try to break the sequence with an object. Particles may follow the object, but prepositions always precede them (Vestergard, 2019, p.4): *She **switched** <u>the light</u> **off***, but not *The pen **fell** <u>the table</u> **off***. In this study, we only consider phrasal verbs. In unclear cases when particles and prepositions were not easily distinguishable, we consulted dictionaries (e.g., Cambridge Online Dictionary, Merriam-Webster.com Dictionary). The dictionaries list phrasal verbs as single lexical units, but they do not have separate entries for verbs with prepositions.

In German, a comparable verbal construction to the phrasal verb is the separable verb – Trennbare Verben (Thim, 2012, pp.3-4). These consist of a verb and a prefix, and similarly to phrasal verbs, the prefix changes the meaning of the verb and separates from it in certain situations, e.g., in conjugated forms: *sagen (say) – absagen (cancel), Sie **sagten** das Meeting **ab** (They cancelled the meeting)*. Not all prefixes in German verbs are separable; if they are not, the verbs are called “inseparable” (Untrennbare Verben). Although English phrasal verbs and German separable verbs are similar, they are not in perfect correspondence. In English, the particle is always separated and follows the verb; in German, there are specific rules for when the prefix precedes or follows the verb (Thim, 2012, p.4). Still, this is another example of discontinuity.

3. METHODOLOGY

As we have stated in the previous sections, our aim for this study is to see how well the model translates from German to English. We know that NMT models are gaining higher and higher BLEU scores, and we know that the state-of-the-art transformers capture some syntax in language. But we do not know how they handle scenarios when the lexical forms are broken. Motivated to look into this matter, we have chosen a specific case of discontinuous units in English – phrasal verbs – and set up a more specific research question: does an NMT model connect the discontinuous parts that belong together? In our specific case, how does it handle phrasal verbs?

To address the formulated questions, we need a collection of sentences that include the pattern under study (phrasal verbs) and a neural machine translation model to translate them. Below we explain in detail which model we use for this study and how we collect the data for translation.[1]

3.1 Facebook FAIR Model

To select a model for this research, we consulted a survey of available pre-trained models by Facebook AI Research included with the fairseq toolkit – an open-source toolkit for sequence modeling (Ott et al., 2019); we chose the newest model from 2019 (Ng et al., 2019). This model

[1] Our data and code are available at https://github.com/LaNataliia/nmt-of-phrasal-verbs.

became a winner of the WMT19 shared news translation task at the Fourth Conference on Machine Translation, and it is available for two language pairs: English-German and English-Russian (in both directions). For the needs of this research, we selected the German-English model – *transformer.wmt19.de-en*.[2]

The Facebook FAIR model from 2019 (Ng et al.) consists of 3 Transformers whose feed-forward network (FFN) layers are of size 8192 (Barrault et al., 2019, p.10). It is a sentence level system trained on the following corpora suggested by the WMT19 committee: Europarl, ParaCrawl, Common Crawl, News Commentary, Wiki Titles, and Document-split Rapid corpus. At the data pre-processing stage, the authors apply the Moses tokenizer (Koehn et al., 2007) for tokenizing the corpora; the problem of rare words is addressed with the subword segmentation method from Sennrich et al. (2015). The model scores 40.8 on BLEU (Ng et al., 2019) and outperforms the other models in WMT19 as per human evaluation (Barrault et al., 2019, p.24).

3.2 TED Talks Corpus

As we work with translation, we need a collection of parallel sentences to test our selected model on. Among the available open-source parallel corpora, we chose to work with the transcribed and translated TED Talks for a number of reasons:

- NMT models are trained mainly on written texts, and TED Talks is an example of spoken language. Asking the model to translate transcribed spoken data, we can see how it can handle a register different from what it was trained on;
- TED Talks are delivered by different speakers, and each speaker has his/her own style of communication. Thus, the texts differ in language usage;

[2] fairseq's pre-trained NMT models are available at https://github.com/pytorch/fairseq/tree/master/examples/translation.

- TED Talks cover a wide range of topics, so the texts also differ in vocabulary.

A large database of TED Talks available for research purposes is distributed with the WIT3 corpus (Cettolo et al., 2012). This corpus is suitable for machine translation tasks as it puts together the original TED public lectures (usually held in English) and their translations into other languages provided by the TED Conference website.[3] In this study, we used a small sample of these TED Talks included with the ParCor dataset (Guillou et al., 2014). ParCor is a parallel English-German corpus annotated for pronoun coreferences; we took the raw documents with no annotation but with sentence alignment.[4] Overall, this collection included 11 TED documents with a total of 1776 parallel sentences (Appendix A).

To prepare these texts for use in our study, we applied basic data pre-processing. The raw documents were split into separate files, so firstly, we combined them into single files (separate for German and English), maintaining the alignment of sentences. This was done by manipulating the files in Python 3.6.9[5], in the Google Colaboratory environment (Bisong, 2019). Secondly, we normalized some particular character combinations: "␣'" (starting with white space) was replaced with the apostrophe character ('), """ was replaced with the single quote character ("), and "&" – with the ampersand character (&). This step was done in Sublime Text 3 editor[6] using its search-and-replace edit option.

3.3 Phrasal Verbs Subcorpus

Since we focus on a specific pattern in language, i.e., phrasal verbs, we do not need to work with the whole collection of TED texts; we only need to look at the sentences that contained phrasal verbs in English. To retrieve this subcorpus, we manipulated our data in Python again,

[3] https://www.ted.com/
[4] Data available at https://opus.nlpl.eu/ParCor/.
[5] Python Software Foundation; available at https://www.python.org/.
[6] Sublime HQ Pty Ltd; available at https://www.sublimetext.com/.

using the spaCy library (Honnibal et al., 2020) for text processing. There are pipelines of different sizes available for English in spaCy (small, medium, large); we used the large pipeline in order to obtain more precise results.

First of all, we took the English TED file, and we searched it for the sentences that contained phrasal verbs. There are two possible approaches to this task: searching via dictionaries using the lists of phrasal verbs or searching via grammatical patterns for verbs with particles. However, the first approach returns too many false positives. Phrasal verb particles have the same form as prepositions. For example, *on* in *turn on the TV* is a particle, but in *turn on Monday*, it is a preposition. Hence, only looking for a dictionary combination *turn on* is not enough to find a phrasal verb in a sentence.

The second approach is based on syntactic parsing. A built-in parser in spaCy is able to distinguish between particles and prepositions, and it assigns different POS tags and dependency labels to them. For example, in the sentence *John is **going on** a trip*, token *on* is parsed as a preposition: its POS-tag is IN (a tag for preposition or subordinate conjunction; Marcus et al., 1993), it is dependent on the verb *going*, and the dependency type between them is "prep," which stands for "preposition" (the visualization of the dependency tree of this sentence is shown in Appendix B, Figure B1). In case of *What is **going on** here?*, token *on* is parsed as a particle: its POS-tag is RP (a tag for particle; Marcus et al., 1993), it is also dependent on the verb *going*, but this time, the dependency type between them is "prt," which stands for "particle" (Appendix B, Figure B2 contains the visualization of the dependency tree of this sentence).

The second approach returns less false positives, so we prioritize it for searching the sentences with phrasal verbs. We used the grammatical patterns and syntactic dependencies observed in the examples discussed above to define our search criteria. In this way, a parsed sentence:

- had to have a particle (a token with the tag IN from the Penn Treebank tagset),
- whose parent was a verb (it depended on a token with the tag VERB from the Universal tagset – using this universal tag was more compact than listing six different tags for various verb forms from Penn Treebank),
- and the relationship between them was "particle" (the dependency label was "prt").

If all three conditions were met, we picked the sentence. Additionally, we recorded its index in the English file. As a result, we collected 141 instances of phrasal verbs in 131 sentences (some sentences had two instances of phrasal verbs – we kept both entries of the same sentences for the convenience of further analysis).

Once we had the English sentences, we used the collected indices to query the German file and retrieve the corresponding German sentences. For accessibility convenience, we saved the results into a table, specifically, a pandas DataFrame (McKinney, 2010).

3.4 Data Translation and Cleaning

When our subcorpus was collected, we were ready to launch the Facebook FAIR model by Ng et al. (2019) and translate the German sentences into English. We added the obtained translations to the DataFrame and exported the result into an Excel[7] file.

To finalize the data collection process, we inspected all our sentences manually and excluded those that were classified incorrectly. Firstly, we removed all false positive instances of phrasal verbs. The precision of spaCy's parser is not 100% and sometimes it does identify prepositions as particles, as in this case: *This is an amazing story and adventure for you to **go on*** (*on* is a preposition here – *to go on an adventure*). Secondly, we removed the sentences where phrasal verbs acted as noun modifiers or appeared as participles that could be classified as adjectival, for instance: *This applies to laypeople thinking about their own happiness, and it*

[7] Microsoft Corporation; available at https://office.microsoft.com/excel.

applies to scholars thinking about happiness, because it turns out we're just as ***messed up*** *as anybody else is*. Additionally, we removed the instances where human translations into German did not correspond to the structure of the original English sentences to an extent where the target verb disappeared. For example, we excluded the pair *You* ***give*** *the knife* ***out*** *– Und hier ist das Messer* (the literal translation of this German sentence is *And here is the knife,* so there would be no reference for the model to generate a phrasal verb out of this sequence). Lastly, we removed the entries with not full sentences in either English or German, such as *That's* ***headed up*** *to about nine billion* (a full sentence) – *die sich auf 9 Milliarden zu bewegen* (a fragment).

While inspecting the sentences for false positives, we annotated the remaining data and added the following information to the DataFrame: the original phrasal verb (the infinitive), the human-translated German counterpart (the infinitive), the corresponding system output (either a phrasal or a regular verb in the infinitive), whether the original target verb goes together with the particle in a sentence, and whether the German counterpart is a verb with the separable prefix. The final version of the data table contained 119 samples.

4. RESULTS

In this section, we describe how we analyzed the collected data, and what results we obtained. It is essential to mention that we did not compare the original English sentences with the system generated English sentences. The original TED Talks were held in English and translated into German by human translators. As we have seen previously, human translators sometimes change the sentence structures, reduce or add additional information. And when we pass this data to the machine translation system, this is not the original text anymore. This is already a modified text and therefore its translation is not comparable to the original. Hence, we used the original English data solely for constructing the subcorpus of the sentences with the phrasal verbs. When we filtered by the original verbs, we knew that their German counterparts chosen by human translators would have equivalents in English that were phrasal verbs. Our focus then was on how the system would handle those cases: would it generate phrasal verbs or would it opt for morphologically simpler, non-phrasal synonymous verbs?

4.1 General Description of the Dataset

The Original English data contained 70 unique phrasal verbs; of them, 47 phrasal verbs occurred only once, and 23 repeated two or more times (maximum 8). Additionally, there were 3 instances of non-phrasal verbs. Initially, those sentences appeared in the dataset because they contained phrasal verbs in other parts and were selected by the filtering algorithm. When inspecting the data manually, we noticed that the corresponding sentences in German had slightly different constructions, and it resulted in the model generating phrasal verbs where they were not present in the original text. For example, the pair *But these stories bothered me, and I couldn't* ***figure out*** *why, and eventually I* ***did*** (original EN) – *Aber diese Artikel machten mir Sorgen, und ich konnte nicht* ***herausfinden****, warum, und schließlich* ***fand*** *ich es* ***heraus*** (DE counterpart); initially, it was selected for *figure out* in the original English sentence, which corresponded to the German verb *herausfinden*. But the German sentence contained this verb two times (in the original, it was substituted with *do* to avoid repetition), so the model generated it two times as well: *But these articles worried me, and I couldn't figure out why, and eventually I* ***found out*** (system EN). We kept this instance as a separate sample in the dataset because, as noted previously, we were looking into how the model generated phrasal or non-phrasal verbs when translating from German.

The corresponding German data contained 75 unique verbs, and 57 of them occurred only once; the other 18 repeated two or more times (maximum 8). Among these 75 verbs, there were 23 verbs that had no prefix, 13 verbs that had an inseparable prefix, and 39 verbs with a separable prefix (including 3 reflexive verbs: *sich daranmachen*, *sich herausstellen*, *sich anmelden*). Additionally, there were cases when the original English phrasal verbs had counterparts in German that were not verbs but rather fixed phrases with or without verbs. For instance, we had 2 cases with no verb at all when the original English *end up* or *wind up* corresponded to *am Ende* in German (e.g., *But now, because the reflective self is in charge, you may* ***end up*** *-- some people may end up moving to California* (original EN) – *Nun aber, weil das nachdenkliche Selbst verantwortlich ist, können Sie* ***am Ende*** *-- einige Leute könnten am Ende nach Kalifornien ziehen* (DE counterpart)). And we had 10 cases with fixed phrases such as *keine Zeit mehr haben*, *sich auf den Suche machen*, *zum Ende kommen* (e.g., *So, I'm going to* ***wrap up*** *now* (original EN) – *Ich* ***komme*** *nun* ***zum Ende*** (DE counterpart)).

In the data generated by the model we had both phrasal and non-phrasal verbs, and non-phrasal verbs were prevailing. In 119 translated sentences, 36 unique phrasal verbs were generated 54 times: 27 non-repeating and 9 repeating phrasal verbs (maximum number of repetitions – 8). In 65 remaining cases, the model produced the following non-phrasal outputs:

- 52 instances of 38 unique non-phrasal verbs; 30 verbs were non-repeating, and 8 verbs repeated two or more times (maximum 5);
- 12 instances of 11 unique fixed phrases (e.g., *come to mind*, *drive crazy*, *in the end*) or constructions like verb + adjective (e.g., *keep awake*), verb + preposition (e.g., *think about*);
- 1 instance of no verb at all as a result of the model reducing half of the sentence (this case will be discussed in detail in 4.4).

The general description of the dataset is summarized in Table 1. The column "Unique" shows the count of all unique entries, including verbs and fixed phrases (e.g., in "System EN Total → Non-phrasal Verbs", 50 unique values include 38 unique non-phrasal verbs, 11 unique fixed phrases, and 1 entry with no verb at all). The column "Top" lists the most frequent entry, and the column "Frequency" shows how many times it appeared (e.g., the most frequent phrasal verb generated by the model is *turn out*, and it appeared 8 times).

Table 1

Summary of the Dataset

	Count	**Unique**	**Top**	**Frequency**
Original Phrasal Verbs	119	73	turn out	8
DE Counterparts	119	86	sich herausstellen	8
System EN Total	119	86		
• *Phrasal Verbs*	54	36	turn out	8
• *Non-phrasal Verbs*	65	50	get	5

4.2 Discontinuity: Correlation Between the Types of English and German Verbs

Discontinuity in our dataset was present in both German and English verbs: in the form of separable prefixes in German, and in the form of phrasal verbs composed of a verb and a particle in English. However, our dataset was not homogeneous. Among the German verbs, we also had inseparable verbs or verbs with no prefixes; they were translated by the model to non-phrasal verbs in English more often than to phrasal verbs.

In this section we present our analysis on the subject of any regularities between the discontinuity in the input (German) and the output (English) as reflected in the model. For the input, we took into account only 107 samples with single token verbs in German and excluded the fixed phrases for this part of the analysis. In the corresponding translated English data, however, some phrases were still present; they were considered as “non-phrasal verb” system output, along with verbs such as *build*, *produce*, *open*, etc.

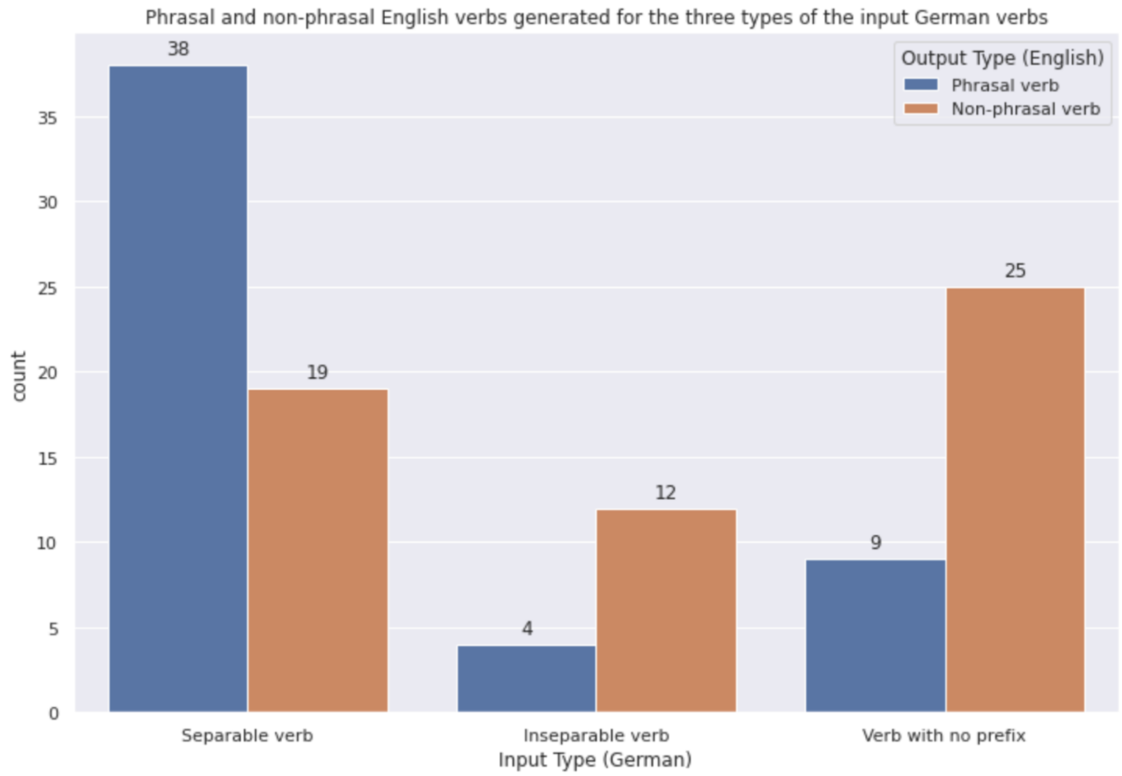
After excluding the German fixed phrases, we had the input data of 3 types and the output data of 2 types:

- input type (German): separable verbs, inseparable verbs, verbs with no prefix;
- output type (English): phrasal verbs, non-phrasal verbs.

We added these additional variables (columns) to the DataFrame, and visualized them in a countplot (Figure 1).

Figure 1

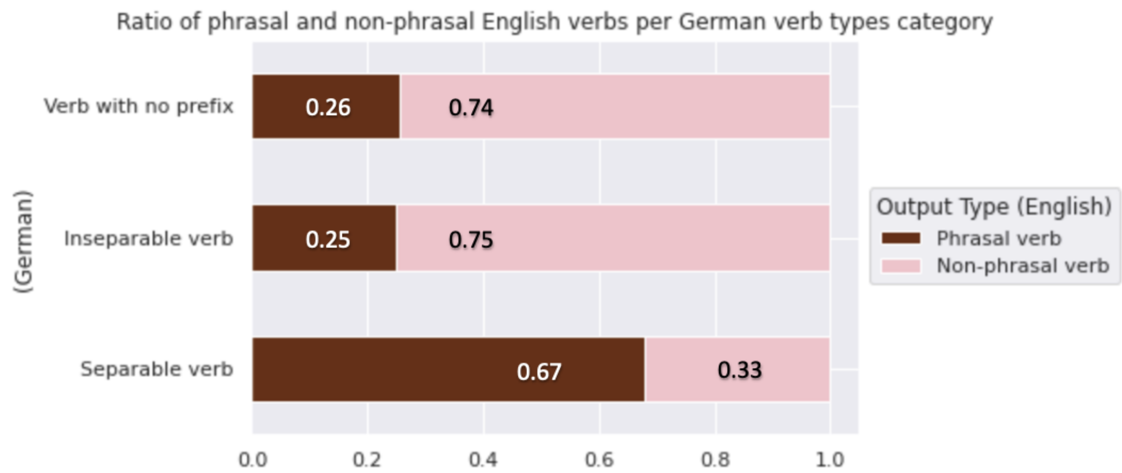
Phrasal and Non-phrasal English Verbs Generated by the Ng et al.’s (2019) Model for the Three Types of the Input German Verbs



As we see in Figure 1, the model generated phrasal verbs more frequently than non-phrasal verbs in English only when translating German separable verbs; in the case of German inseparable verbs or verbs with no prefix, the model generated non-phrasal verb translations more frequently. We proceeded with plotting the ratio of phrasal and non-phrasal output verbs per German input type category (considering each category as 1.0 or 100%), and we got the result as shown in Figure 2.

Figure 2

Ratio of Phrasal and Non-phrasal English Verbs Generated by the Ng et al.'s (2019) Model for the Three Types of the Input German Verbs



The pattern that we observed on these visualizations implied a certain relation between the discontinuity in German and English verbs as perceived by the Ng et al.'s (2019) model. We could establish whether this relation was statistically significant or not by applying a statistical test for independence. As we had two categorical variables "Input Type (German)" and "Output Type (English)" that consisted of three and two independent groups (Separable verb/ Inseparable verb/ Verb with no prefix and Phrasal verb/ Non-phrasal verb, respectively), an appropriate test choice in this case was the chi-square test for independence (Oakes, 1998, p.25).

To test whether the two variables were independent of one another, we started with formulating the null hypothesis (H_0). In our case, the H_0 stated that there was no relation between the type of German verb the model received as an input and the type of English verb the model generated as an output. The alternative hypothesis (H_a) stated that there was a relation, and it was statistically significant.

To run the test, we prepared a contingency table for the two variables (Appendix C, Table C1), but it contained one value below 5, and this number was too small to run the chi-square test (Oakes, 1998, p.25). As both inseparable verbs and verbs with no prefix in German are single units with no separable parts, we could combine these two groups under the joint characteristic "Non-separable verbs" and therefore address the issue of low frequencies. The new contingency table (Appendix C, Table C2) did not contain values below 5. Thus, we could run the chi-square test.

We ran the test in Python using the SciPy toolkit (Virtanen et al., 2020), and we obtained the following values:

- chi-square value (χ^2) – 16.066,
- p-value – 6.1184e-05,
- degrees of freedom – 1 (for 2x2 contingency table).

We reject the null hypothesis when the calculated χ^2 value is equal to or greater than the critical value (Oakes, 1998, p.25). The critical value for chi-square at 0.05 level of significance (the alpha value generally chosen in linguistics) and 1 degree of freedom is 3.84 (Oakes, 1998, p.266). The calculated χ^2 value (16.066) was greater than that, so we rejected the null hypothesis. Additionally, the p-value was not greater than the alpha value (0.05), and it showed once again that the null hypothesis did not hold true. In this way, we established a statistically significant relation between the type of German verb the model received as an input and the type of English verb the model generated as an output. This meant the model was more likely to generate a phrasal verb in English when translating a separable verb from German.

Next, we could look at the model behavior when the prefix of a German verb was actually separated in an input sentence. We ran another chi-square test, taking into account only 57 samples with separable verbs in German. We annotated each sample for whether the prefix was separated from the verb in the German sentence, adding the variable "Input: Is Separated?" with the values "Yes" and "No" to the DataFrame. The "No" category included the cases with or without an infix (*zu* or *ge*) after the prefix – even though an infix was inserted, the prefix still stayed within the token boundaries, and the verb formed one token with the prefix; the "Yes" category included the cases when the verb and the prefix formed two tokens. The null hypothesis we wanted to test here stated that there was no relation between the two characteristics: the model received a German verb with a prefix separated from it, and the model translated it into a phrasal verb in English. The alternative hypothesis stated that the two characteristics were dependent, and the model was more likely to generate a phrasal verb in English when it received a German verb with a prefix separated from it.

To test the null hypothesis, we performed the same actions as before: we prepared a contingency table (Appendix C, Table C3), we ran the chi-square test for independence in SciPy, and we compared the calculated χ^2 value with the critical value. As we had the same degree of freedom (1) and alpha value (0.05), the critical value also stayed the same – 3.84. After running the test, we obtain the χ^2 value 0.080966 (not greater than or equal to the critical value) and the p-value 0.77599 (greater than the alpha value). Such values meant that the null hypothesis held true, and there was no statistically significant relationship between the two variables. In other words, the model was able to generate a phrasal verb independently of whether it received a separable German verb in one or two tokens.

4.3 Consistency of the Model

So far, we have been looking only into the numbers and the quantitative characteristics of our dataset. In this section we present our analysis of the actual translations and of whether the model was consistent in what it produced. We did the analysis separately for the repeating and non-repeating input (German) verbs. The main resources we consulted for this part were Langenscheidt Online Dictionary, Cambridge Online Dictionary, Collins Online Dictionary, Merriam-Webster.com Dictionary, Google Books Ngram Viewer, and Google Translate with its frequency information – the scale that indicates the frequencies with which certain translations of a given word or phrase appear in public documents.

4.3.1 Repeating verbs

In the German data, we had 18 verbs that repeated two or more times. We could perform a qualitative analysis of how the model translated them into English – was the output the same across the cases, or was it different? We went verb by verb, starting with the most frequent ones.

sich herausstellen. This separable reflexive verb was the most frequent among the German verbs in our dataset; it appeared 8 times. All 8 times, the model translated it in the same way: *turn out*. This translation was the only one listed in Langenscheidt Online Dictionary; though, Cambridge Online Dictionary gave another option – *transpire*. In one case, the model changed the tense from past to present: *Es* ***stellte sich*** *aber* ***heraus****, dass das noch nicht alles ist* ⇒ *But it* ***turns out*** *that's not all* (original index 375). In one more case, the model changed the tense and also the structure of the sentence: (369) *Es* ***hat sich*** *nämlich Folgendes* ***herausgestellt****: Wenn Sie ein Placebo in Form einer weißen Pille anbieten, in der Form einer Aspirintablette, ist das einfach eine runde, weiße Pille, die einen bestimmten messbaren Effekt hat* ⇒ *It* ***turns out*** *that if you offer a placebo in the form of a white pill, in the form of an aspirin* *tablet, it is simply a round, white pill that has a certain measurable effect*. The literal translation of the German construction would be similar to *namely, it turned out to be the following…* But the model simplified the structure by reducing some words without losing or changing the meaning of the sentence.

herausfinden. It appeared 5 times, and it was translated in 2 different ways: *figure out* and *find out*. In the translated sentences, *figure out* appeared before the interrogative adverbials (e.g., (649) *Und Sie müssen* ***herausfinden****, wie Sie Ihr echtes Leben gestalten, wenn es wahr wäre ⇒ And you have to* ***figure out*** *how to shape your real life if it were true*), while *find out* went before *that*-clause (e.g., (1544) *Sie nehmen das Zeug, und sie nehmen alternative Heilmittel, und es macht keinen Unterschied, wie häufig wir* ***herausfinden****, dass sie nutzlos sind ⇒ They take the stuff, and they take alternative remedies, and it doesn't matter how often we* ***find out*** *that they're useless*). In both examples, these English phrasal verbs were not interchangeable, so the model seemed to adjust to the context and generate a translation that fitted best. In another example, *herausfinden* was translated in two different ways within one sentence: (1460) *Aber diese Artikel machten mir Sorgen, und ich konnte nicht* ***herausfinden****, warum, und schließlich* ***fand*** *ich es* ***heraus*** *⇒ But these articles worried me, and I couldn't* ***figure out*** *why, and eventually I* ***found out***. We noticed that in the first case, *figure out* appeared before the interrogative adverb again, while *find out* had no object or clause. A similar German construction was found in (1449), but there the model did generate an object: *Und ich habe eine Weile gebraucht, um das* ***herauszufinden*** *⇒ And it took me a while to* ***figure*** *that* ***out****.*

emittieren. This was a verb with no prefix that appeared 4 times in the dataset. In each case, the model generated the same non-phrasal verb translation – *emit*. When we looked up this German verb in Google Translate, *emit* appeared to be its most frequent translation in public documents (Appendix D, Figure D1; the frequency scale is on the right). Besides being the most frequent translation, it was the only one that fitted the context of our sentences because all of them talked about CO2 emission: (748) *Fast jede Herstellungsmethode für Strom* ***emittiert*** *heutzutage CO2, außer die erneuerbaren Energien und Nuklearenergie ⇒ Almost every method of producing electricity nowadays* ***emits*** *CO2, except for renewables and nuclear energy*. The phrasal verb options were not listed as possible translations there. All of the original English sentences contained the phrasal verb *put out*, e.g., (748) *Almost every way we make electricity today, except for the emerging renewables and nuclear, puts out CO2* (original EN). But here, the issue was that the dictionaries did not translate *put out* as *emittieren* into German. The human translator had chosen this translation because it was the best one to fit the context. However, we could not expect the model to generate *put out* in those cases because it was not a listed translation of German *emittieren*.

steigen. It occurred 4 times, 2 times it was translated as *rise*, and 2 other times as *increase*. In the original English data, those sentences all contained the phrasal verb *go up*, e.g., (735) *And it's a great thing for this number to go up* (original EN). When we consulted Google Translate, *go up* appeared as a less common translation for *steigen*, while both *rise* and *increase* appeared as common (i.e., frequent; Appendix D, Figure D2). So the model seemed to opt for more common, non-phrasal verb translations. But when did it differentiate between *rise* and *increase*? If we compare these two verbs, *rise* is more neutral, as it can be used in many different situations: *the sun rises, taxes rise*, *rates rise*, *rise to the sky*, *rise from the chair*, etc. *Increase* is more specific: rates and prices can increase, but we cannot say *the sun increases* or *increase from the chair*. In our examples, we noticed that the neutral *rise* appeared in short sentences: (735) *Es ist toll, dass diese Nummer* ***steigt*** *⇒ It's great that this number is* ***rising***. In longer sentences, the model generated the more specific *increase*: (736) *In der reichen Welt, in der oberen Milliarde, könnten wir wohl Abstriche machen und weniger nutzen, aber im Durchschnitt wird diese Zahl jedes Jahr* ***steigen*** *und sich somit insgesamt mehr als verdoppeln, die Zahl der Dienste die pro Person bereitgestellt werden ⇒ In the rich world, in the top billion,* *we could probably cut back and use less, but on average this* *__number__* *will* ***increase*** *every year and thus more than double the total number of services provided per person*. What was peculiar, however, is that in both translations, our target verb had the same subject in English – *number*. But in the German sentences, 2 different nouns were used: *Nummer* and *Zahl*. In German, they have different meanings and are not interchangeable in every context: every *Zahl* is a *Nummer*, but vice versa is not true (e.g., in *Nummer drei im Mobilfunkmarkt ist die Firma Salt*, noun *Zahl* cannot be used instead of *Nummer*). So in those cases, the model took into account not only the context of what it produced but also the context of what it received as an input. As a result, we got an English verb with a broader meaning in the context of a German noun with a broader meaning, and a more specific English verb for a more specific noun in German. One of the explanations of this choice of the model could have been stylistic.

__aufgeben__. There were 3 entries of this verb in the dataset, twice it was translated as *give up*, and one time as *abandon*. Here the difference was in the transitivity of the German verbs. In case of intransitive verbs with no object, the model translated it as a phrasal verb: (1) *Ich habe einige Leute die Anzahl der Bücher zählen lassen, die mit "Happiness" im Titel in den letzten fünf Jahren veröffentlicht wurden und sie* ***gaben*** *nach ungefähr 40* ***auf****, und es gab noch viel mehr* ⇒ *I had some people count the number of books published with "Happiness" in the title in the last five years and they* ***gave up*** *after about 40, and there were many more*. For a transitive verb with a direct object, the model generated a non-phrasal verb translation: (770) *Diese leisten vielleicht einen moderaten Beitrag und wenn sie besseres leisten als ich erwarte, wäre das toll, aber meine Kernaussage hier ist, dass wir an all jenen fünf arbeiten müssen, und wir können* *__keine von ihnen__* ***aufgeben****, weil sie uns einschüchtern, denn alle haben signifikante Probleme* ⇒ *They may make a moderate contribution, and if they do better than I expect, that would be great, but my core message here is that we have to work on all those five, and we cannot* ***abandon*** *__any of them__* *because they intimidate us, because they all have significant* *problems*. Even though the original English sentences all contained *give up*, with or without direct object, the model differentiated these cases.

The remaining 13 verbs repeated only two times each. 9 of them, the model translated in the same way:

- ***aufwachsen*** ⇒ *grow up* (the only possible translation);
- ***bekommen*** ⇒ *get* (the most frequent translation according to Google Translate frequency scale, the original phrasal verbs *tease out* and *get up* were not among the listed translations);
- ***finden*** ⇒ *find* (the most frequent translation according to Google Translate frequency scale, the original phrasal verb *come up* was not among the listed translations);
- ***gelangen*** ⇒ *get*

According to Google Translate, the most frequent translation of *gelangen* was *reach*. However, in both German sentences, *gelangen* appeared with a preposition (*an* or *zu*). It resulted in the translation *get to somewhere*, e.g., (1575) *Sie* ***gelangen*** *zu Thabo Mbeki in Südafrika* ⇒ *You* ***get*** *to Thabo Mbeki in South Africa* – as opposed to *reach*, which does not require a preposition. The original phrasal verb *end up* was not among the listed translations;

- ***herausnehmen*** ⇒ *take out*

Other translations of this verb such as *take away*, *pull out*, *remove*, or *extract* were also possible, and they were of similar frequency, according to Google Translate. But the model opted for a phrasal verb translation of a separable German verb. For example: (431) *Normalerweise würde ich jetzt die Nadel* **herausnehmen** ⇒ *Normally I would* **take** *the needle* ***out*** *now*. This translation also coincided with the original phrasal verb;

- ***öffnen*** ⇒ *open* (the most frequent translation according to Google Translate frequency scale; the original phrasal verb *open up* was not among the listed translations);
- ***passieren*** ⇒ *happen*

It was one of the most frequent translations, according to Google Translate, along with *pass*. But *pass* would not fit the context in our examples: (1640) *Es gibt keine andere Beschreibung für das, was hier* ***passiert*** ⇒ *There is no other description of what is* ***happening*** *here*. In regard to the original phrasal verbs, *show up* was not among the possible translations, and *go on* was a rare translation, according to Google Translate;

- ***vermasseln*** ⇒ *screw up*

It was one of the most frequent translations according to Google Translate, along with *mess up* (which was an original phrasal verb). Either option would fit the context in our sentence that contained both occurrences of this verb: (118) *Jeder, der diese Begriffe nicht unterscheidet wird die Erforschung des Glücks* ***vermasseln****, und ich gehöre zu einer Menge von Forschern über Wohlbefinden, die die Erforschung des Glücks lange* ***vermasselt haben*** *auf genau diese Art* ⇒ *Anyone who does not distinguish these terms will* ***screw up*** *the study of happiness, and I am one of a lot of researchers on well-being who have long* ***screwed up*** *the study of happiness in exactly this way*;

- ***wachhalten*** ⇒ *keep awake*

The dictionaries provided *keep alive* or *keep up* as the translations of this verb, and the latter was an original phrasal verb. In our sentence, which contained both occurrences of this verb, it co-occurred with the adverb *nachts* (*at night*): (1440) *Aber das hier* ***hält*** *mich nachts* ***wach*** -- *eines der Dinge , die mich nachts* ***wach halten*** ⇒ *But this one* ***keeps*** *me* ***awake*** *at night* -- *one of the things that* ***keeps*** *me* ***awake*** *at night*. Thus, the model seems to infer the meaning from the context and generate the adjective *awake* instead of *alive* or *up*.

For 4 other German verbs that repeated twice in the dataset, the model generated different translations. But in all cases, the translations fitted the context:

<u>aufbauen</u>, translated as *build* and *set up*. Though *set up* was a less frequent translation than *build up*, the model adjusted to the context when needed: (455) *Also hat er vor den Spielern <u>eine Kamera</u>* ***aufgebaut*** ⇒ *So he* **set up** *<u>a camera</u> in front of the players*.

<u>gehen.</u> Its most frequent translation into English is *go*; however, the model did not generate this translation in either case. In (832), the model translated it as *get*: *Nun,* ***gehen*** *wir <u>bis ans Ziel</u>, das wir erreichen müssen und dann reden wir über den Zwischenschritt* ⇒ *Well, let's* ***get*** *<u>to the goal</u> we need to achieve and then we'll talk about the intermediate step*. Dictionaries do not provide *get* as a translation for *gehen*. But when we consulted the Google Books Ngram Viewer, we saw that the phrase *get to the goal* in English was much more frequent than *go the goal*, especially in recent years (Appendix E, Figure E1). The model seemed to take it into account.

In (432), we had a similar case: *Ich würde meinen Arm säubern und Ihnen zeigen, dass da keine Wunden sind. Aber ich denke, in diesem Rahmen hier und mit der Absicht, aus einer Täuschung etwas Echtes zu machen, werde ich die Nadel einfach da drin lassen und so <u>von der Bühne</u>* ***gehen*** ⇒ *I would clean my arm and show you that there are no wounds, but I think in this context here and with the intention of turning a deception into something real, I will just leave the needle in there and* ***walk off*** *<u>the stage</u>*. The literal translation of the phrase *von der Bühne gehen* would be *go off the stage*; the model translated it as *walk off the stage*. When we compared these two phrases through the Google Books Ngram Viewer, we saw that the latter was more common and frequently used (Appendix E, Figure E2).

<u>kommen</u>. In one case, the model produced an expected translation *come*, which was also the most frequent as shown in Google Translate: (757) *Normalerweise steht man quasi nur daneben und manche* ***kommen****, andere nicht* ⇒ *Normally you just stand there and some* ***come*** *and some don't*. In the original sentence, it was *come along*, but this translation was quite rare, according to Google Translate. In the other example, the translation coincided with the original phrasal verb: (1322) *Zu mir* ***kommen*** *junge Filmemacher, die sagen: "Geben Sie mir einen Rat, wie ich so etwas machen kann."* ⇒ *Young filmmakers* **come up** *to me and say, "Give me some advice on how to do this."* In the dictionaries, *come up* did not appear as a translation for *kommen*, but it might have been influenced by preposition *zu*: *come up to someone* means to approach at a physical gathering, where initially both parties are present; *come* implies that initially the parties are not in the same location. *Zu mir* indicated the physical presence, so *come up* is a relevant choice.

<u>tun</u>. By itself, this verb stands for *do* in English, and in one of the sentences it was translated exactly this way: (559) *Der Grund ist, dass, wenn wir spielen, sind wir tatsächlich glücklicher dabei, hart zu arbeiten, als wenn wir uns entspannen oder <u>nichts</u>* ***tun*** ⇒ *The reason is that when we play, we are actually happier working hard than when we relax or* **do** *<u>nothing</u>*. In the original sentence, the phrasal verb was *hang out*; German *tun* does not really have this translation in English, even though *nichts tun* has a similar meaning. In the other sentence, tun co-occurred with the phrase *am Ende*, and all together, the model translated it as *end up doing*; this translation coincided with the original phrasal verb: (134) *Wenn Sie also das Glück der beiden Arten des Selbst maximieren wollen werden Sie <u>am Ende</u> sehr unterschiedliche Dinge* ***tun*** ⇒ *So if you want to maximize the happiness of the two types of self, you will <u>end up</u>* ***doing*** *very different things*.

Am Ende appeared in two more sentences in our dataset. When it appeared in combination with another verb (or in relation to the same verb even when it was omitted in the input sentence to avoid repetition), the model also translated it as a phrasal verb *end up (doing something)*: (142) *Nun aber, weil das nachdenkliche Selbst verantwortlich ist, können Sie* ***am Ende*** *-- einige Leute könnten* ***am Ende*** *nach Kalifornien* ***ziehen*** ⇒ *But now that the thoughtful self is in charge, you may* ***end up*** *-- some people might* ***end up moving*** *to California*. In the other case, when it just appeared on its own, the model translated it literally – *in the end*.

4.3.2 Non-repeating Verbs and Phrases

In the case of non-repeating German verbs (57 in total), it was not possible to evaluate the consistency of the model in whether it generated the same translations for the different instances of the same verbs. However, we noticed another consistent pattern in its behavior: in more than half of the cases (56% in respect to the total number of non-repeating verbs), it just produced those translations that were more frequent. For 19 verbs, the model generated the most common translation (we considered the most common translation to be the one that appeared first either in the dictionaries or on the Google Translate frequency scale); for 13 other verbs, it generated one of the several common translations.

Only in 10 cases, the model output coincided with the original English phrasal verb. For 15 cases, the original phrasal verb was a listed translation for the given German verb (it appeared in the dictionaries), but it was either of a similar or lower frequency than the model output. Only in one case, the original option (*try out*) was more frequent than how the model translated it (*try*): (1429) *Und aus meiner Sicht ist die wissenschaftliche Methode, Sachen **auszuprobieren**, schauen, ob es funktioniert, es ändern, wenn es das nicht tut, eine der großartigsten Errungenschaften der Menschheit* ⇒ *And from my point of view, the scientific method of **trying** things, seeing if it works, changing it if it doesn't, is one of the greatest achievements of mankind*. In 26 cases, the original phrasal verb did not appear among possible translations in the dictionaries due to the human factor. The human translators did not always follow the original English sentence precisely when translating into German. In some cases, it led to situations when the chosen verb in German did not have the same meaning as the original phrasal verb. Hence, we could not expect the model to generate a translation that would coincide with the original. The following example illustrates this scenario: (767) *Now, there's all sorts gimmicky solutions like that one, but they don't really **add up** to much* (original EN) ⇒ *Nun, es gibt alle möglichen Spielereilösungen wie diese, aber sie **bringen** alle nicht viel* (DE counterpart) ⇒ *Well, there are all sorts of gimmicks like this, but none of them* ***do*** *much good* (system EN). In this case, even if we disregard the original English sentence, the model does not follow the German sentence structure anyway.

There were 13 cases when the model generated a phrasal verb, but it did not coincide with the original due to either the human factor described above or to the model adjusting to the context. For example: (192) *So, I immediately went to* ***look up*** *the 2009 online edition, expecting to find a revision worth noting* (original EN) ⇒ *Also* ***hab*** *ich mir sofort die Online-Ausgabe von 2009* ***aufgerufen*** *und erwartet, dass ich hier einen ansprechenderen Beitrag finde* (DE counterpart) ⇒ *So I immediately* ***called up*** *the online edition of 2009 and expected to find a more appealing contribution here*. In the dictionaries, one of the definitions of *call up* is "to retrieve from the memory of a computer especially for display and user interaction" (Merriam-Webster.com Dictionary), and this translation for *aufrufen* is marked as specific to the computational domain.

With the fixed phrases, the model was relatively consistent: it translated them literally or to the equivalent phrases in English, where possible. For instance:

- (406) ***Macht*** *Sie das* ***verrückt****?* ⇒ *Does that* ***drive*** *you* ***crazy****?* – an equivalent phrase;
- (518) *Aber um es mal im Kontext zu sehen: Vor 5,93 Millionen Jahren* ***begannen*** *unsere ersten Primatenvorfahren* ***aufrecht zu gehen*** ⇒ *But to put it in context, 5.93 million years ago, our first primate ancestors* ***began to walk upright*** – a literal translation. Originally, the whole phrase in the original English sentence was encompassed in the phrasal verb *stand up* (*But to put that in context: 5.93 million years ago was when our earliest primate human ancestors* ***stood up***).

4.4 Particular Cases and Other Observations

While analyzing the translation patterns of the verbs, we noticed a few other peculiarities of the model. They were not limited to the verb choices, but they are important to be mentioned as well.

Simplification. In a few sentences, whether short or long, the model demonstrated the ability to reduce or simplify the sentence it generated compared to what it received as an input. In (374) *Das ließ mich wirklich **ausflippen*** ⇒ *It really **freaked** me **out*** (which also exemplifies the separation of verb and particle), the model dropped out the German verb *lassen*. The literal translation would be *It really made me freak out*. In the output sentence, the general meaning was not changed, but the whole sentence was simplified. A similar case was in (287): *Und Dr. Kean **erzählte weiter**, er **sagte**: "…"* ⇒ *Dr Kean **went on to say**: "…"* The literal translation would be *Dr. Kean narrated further on, he said: "…"*, but the model just combined two verbs with similar meanings (*erzählen, sagen*) into one (*say*). The most striking example of simplification was in (619): *Sie verließen Lydien und **machten sich auf die Suche nach einer neuen Heimat*** ⇒ *They left Lydia **in search of a new home***; the whole predicate was just reduced to a prepositional phrase.

Separation of verb and particle. There were 4 instances when the model generated a phrasal verb with a particle not immediately following the verb; there were one or two words in between them. In 3 cases, the corresponding German verbs were separable (but not separated); in 1 case, it was a verb with no prefix in German:

- (431) *Normalerweise würde ich jetzt die Nadel **herausnehmen*** ⇒ *Normally I would **take** the needle **out** now* (separable verb);
- (719) *Und irgendwie müssen wir Veränderungen hervorbringen, die das auf Null **senken*** ⇒ *And somehow we have to bring about changes that **bring** that **down** to zero* (verb with no prefix).

Verbs that are not in the dictionaries. Our dataset contained 3 German verbs that were present in none of the dictionaries we consulted: *rausgehen*, *reinzoomen*, *rauszoomen*. Nevertheless, the model generated adequate translations for them. For instance: (1355) *Und so bekommt man eine Idee davon, dass wenn man auf diese Art Informationen durchsucht, gezielter, breiter,* ***reinzoomen***, ***rauszoomen***, *dann sucht man nicht oder surft nicht einfach nur* ⇒ *And so you get an idea that if you search information in this way, more targeted, wider,* ***zoom in***, ***zoom out***, *then you're not looking or just surfing*. Google Translate also generated translations for all these verbs. It means that machine translation systems are able to learn more colloquial language as well.

Change in meaning. In most cases, the model produced accurate translations. Only in two cases, the meaning of the German source sentence was changed:

- (1412) *Meine Eltern* ***sind*** *der 80* ***auf den Fersen*** ⇒ *My parents* ***are*** *in their 80s*. The German phrase *auf den Fersen sein* means *to be about to do something*, *to get nearer*. But it does not mean to be there already, as the translated sentence implied;
- (1011) ***Sie*** *hat uns gewissermaßen verkuppelt* ⇒ ***It*** *kind of coupled us together*. This was an example of the classic mistake of the machine translation systems – incorrect choice of the pronoun. *Sie*, however, does not have the translation *it*; it can be either *she* or *you* (singular formal or plural).

5. DISCUSSION

We have carried out an extensive analysis of our data using various tools and resources and applying different quantitative and qualitative techniques. In this section, we will discuss the main findings of our analysis, the limitations of our research design, and the suggestions for further research.

5.1 Main Findings

Overall, the Facebook FAIR model by Ng et al. (2019) demonstrated a high quality of translation. All 119 sentences translated by the model were grammatically correct; 117 sentences preserved the meaning, and only in 2 cases we saw a change in meaning. Our primary focus in the analysis was on the phrasal verbs and their discontinuous nature. The fact that a phrasal verb consists of 2 units, a verb, and a particle, might have been challenging for the model. Nevertheless, the model showed the ability to produce the phrasal verbs, and in few cases, to even separate a particle from a verb by inserting other tokens between them.

We started our analysis by testing whether the discontinuity in the input affects the discontinuity in the output. The analog of English phrasal verbs in German are the separable verbs (Trennbare Verben) that have separable prefixes. In some cases, the prefix stays within the verb but separates from the stem when an infix is inserted (typically, *zu* for the infinitive or *ge* for the past participle); in other cases, the prefix gets completely disconnected from the stem, and the verb is divided into 2 tokens. With the help of the chi-square test for independence, we saw that the model was more likely to produce English phrasal verbs when it translated separable German verbs than when it translated inseparable German verbs or verbs that had no prefix at all. At that, it did not matter for the model whether the prefix was actually separated from the stem or not to produce a phrasal verb in English; those two characteristics were not proved to be correlated.

The peculiarity of our data is that it was collected based on the English original TED Talks. We collected the sentences with phrasal verbs in the original version, knowing that it was possible for those contexts to have phrasal verbs. And then, we only operated with the

German data that was translated into German by people. We aimed to see whether the model would also produce phrasal verbs in those contexts. But we saw that actually, we should not have always expected a phrasal verb due to what we had in the German sentence. Because of the differences in language systems, human translators may change some structures and not follow the original literally to deliver the meaning better. It may lead to situations where a German verb does not have the original phrasal verb translation into English anymore. A good example is the sentences about CO2 emission (715, 724, 744, 748, 927). In all 5 cases, the original construction was *to put out CO2/ carbon dioxide*. Four times it was translated into German as *CO2 emittieren*, and one time as *Kohlenstoffdioxid ausstoßen*. *Put out* is not a specific phrasal verb used mainly in the context of CO2, but *emittieren* is. As a result, German *emittieren* did not have this phrasal translation *put out* listed anymore. But *ausstoßen* did. Both German verbs fitted the context and delivered the meaning correctly, but in 4 cases out of 5, the original phrasal verb was not a listed option for translation anymore. We had around 35 such cases where German verbs did not have phrasal translations into English. When we did not take them into account, we saw that the model produced phrasal verbs in English for the remaining sentences in more than half of the cases (around 60%), even if initially it had seemed that the model opted for morphologically simpler translations more often.

Another example of when we might not expect a phrasal verb from the model was when the particle did not significantly change the meaning of the original verb. For example, in *I would clean off my arm…* (432), particle *off* did not change the meaning of *clean*; it just intensified it. The German equivalent of this phrase was *Ich würde meinen Arm säubern…*, and the model translated it as *I would clean my arm*, without the particle. This was quite expected behavior because *säubern* is simply *to clean*.

In the qualitative analysis, we discovered important features of the model. Namely, the model has some notion of grammar, syntax, and semantics, and it adapts to different scenarios:

1) it can change the verb tenses of the source sentence (sentences 369, 375 with *sich herausstellen*);

2) it can discriminate between transitive and intransitive verbs and translate the source verb differently depending on whether it has an object or not (sentences 1, 10, 770 with *aufgeben*: *we will have to* ***give up***, but *we cannot* ***abandon*** <u>*any of them*</u>);

3) it can simplify constructions without losing the meaning (omitting the verb *lassen* in sentence 374; turning the predicate *machten sich auf die Suche nach einer neuen Heimat* into a prepositional phrase *in search of a new home* in sentence 619; substituting a longer phrase *die Schülern, die* ***die Schule geschmissen hatten*** with a compact phrasal verb: *the students who had* ***dropped out*** in sentence 309);

4) it uses up-to-date language. The model produces phrases and collocations that are more common in language, even if the source verb does not actually have that translation on its own (sentences 432, 832 with *gehen*: get to the goal, walk off the stage). Besides, it copes with the verbs that are not in the dictionaries (sentences 1338, 1355 with *rausgehen*, *reinzoomen*, *rauszoomen*);

5) it is guided by the context.

Regarding the context, we want to highlight several findings. Firstly, the model considers not only the verb itself but also its complements (specific prepositions, certain types of clause, subjects), and not only in the source sentence but also in what it produces. For example, German *gelangen* on its own is *reach*, *attain*, but in sentence 1574, ***gelangen*** <u>*an*</u> *einen Ort* was translated as ***get*** <u>*to*</u> *a place*. The verb *herausfinden* was translated differently when it appeared before the clause that started with *that* or with an interrogative adverb (*we* ***find out*** <u>*that*</u> *they're useless* in 1544; ***figure out*** <u>*how*</u> *to shape your real life* in 649). <u>*Die Zahl*</u> ***steigen*** was *the number* ***increases*** in 736, but <u>*die Nummer*</u> ***steigen*** was *the number is* ***rising*** in 735. Secondly, the model just seems to differentiate the meaning on the level of the whole sentence.

In the context of physical exercises in 209, *aufbauen* was *to* ***build*** *my leg* <u>*muscles*</u>; in the context of shooting in 455, *aufbauen* was *to* ***set up*** *a* <u>*camera*</u> *in front of the players*. The length of the context might also have an impact. In the previous example of *rise/ increase* as the translations of *steigen*, the model generated a more specific verb *increase* when the sentence was long (736), as compared to a more neutral *rise* when the sentence was shorter (735).

5.2 Limitations of the Study

The research methodology designed for this study has shown a number of shortcomings. To begin with, we used spaCy to parse and filter the entire corpus in order to build a subcorpus of sentences with the phrasal verbs. The spaCy parser is not 100% accurate (according to spaCy documentation, it is 0.97 for the POS-tagging and 0.90 for labeled dependencies). We did not have a large subcorpus, so we could go through it and remove the false positive samples. But for larger corpora, it would have been more problematic. Also, spaCy may have missed some false negative sentences. Although it does not influence the quality of the retrieved data, it reduces its quantity.

And the quantity has become another pitfall for the quantitative analysis. Our subcorpus was not large enough to break it into parts and perform cross-validation on several random samplings in the chi-square test for independence.

Lastly, we did not have human translations of our data from German to English. It might have been insightful to compare human translation and machine translation of the same material.

5.3 Future work

In the course of this study, we have seen that the Facebook FAIR model by Ng et al. (2019), as an example of a neural machine translation model with the Transformer architecture, is able to generate discontinuous units such as phrasal verbs. We have also seen on some examples that it takes into account separable prefixes of the corresponding German verbs, prepositions that follow them, whether they have direct objects or not, etc. For the next step in this research, it would be relevant to investigate whether this is a consistent pattern in the model behavior. It can be achieved by looking into the inner mechanisms of the model performance, namely, into the attention patterns. The Transformers have this distinctive feature of self-attention mechanism that improves their performance in various aspects (including accuracy, efficiency, computational costs) in comparison to the previous state-of-the-art architectures. We can look into the attention patterns with the question of what the model attends to when it deals with the discontinuous units when it processes the input sentences and generates its translation.

Based on the limitations of the present study, it would be relevant also to consider resources that have parallel corpora with the human translations into the target language. And since this time we have been working with transcribed spoken data, it could be interesting to include corpora of written texts for the sake of comparison.

6. CONCLUSION

Machines and humans do not "speak" the same language, so it is hard for machines to mimic human language. It is even harder for machines to be multilingual and "speak" several natural languages. Language structure is one challenge for machine translation, but another challenge is language ambiguity, implied meaning, common background knowledge. One more pitfall is the difference between language systems and the realities of the communities where these languages are spoken. Often it is impossible to do a literal word-by-word translation, and some adjustments have to be made. Professional human translators possess this knowledge and information, and they see the whole context of what they translate. Machines only have a corpus to train on and no background knowledge. Nevertheless, machine translation continues to develop and demonstrates good results.

In this study, we aimed to see how well one of the best available pre-trained NMT models translates from German to English. Our specific focus was on the model's ability to handle discontinuous units such as phrasal verbs. Our results show that generally, the model is able to generate phrasal verbs, but it does not always choose to do so. It is more likely to produce phrasal verbs when the input construction is similar, that is, when it translates a separable German verb. In other cases, it depends on the frequency of the phrasal verb in the language in general. If among possible translations of the source German verbs there are highly salient phrasal verb options that are frequent in English (e.g., *turn out*, *give up*, *figure out*), the model has no problem generating them. If there are other morphologically simpler options of similar frequency, the model is more likely to opt for them. Regardless, the model has demonstrated that it can make a connection between two discontinuous units of one lexical item. In few cases, it even showed the ability to interrupt this item and place an object between the verb and the particle. We also note that, in general, the overall quality of the translation was high. The model generates grammatically correct sentences that correspond semantically to the source sentences. Moreover, it demonstrates an ability to learn certain syntactic patterns, capture semantics, and adapt to the context.

The novelty of this study consists in analyzing the discontinuous units in neural machine translation. Our results contribute to a better understanding of how much linguistic knowledge one NMT model has in general and how well it can handle discontinuity in particular. This could be useful for NLP engineers in further development of such models. Finally, the practical value of this paper consists in looking beyond the automatic BLEU score and evaluating neural machine translation qualitatively. Moreover, we suggest a practical method of making a subcorpus with a specific pattern of interest that can be adapted to other needs and used in further research.

REFERENCES

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv preprint arXiv:1409.0473.*

Barrault, L., Bojar, O., Costa-jussà, M. R., Federmann, C., Fishel, M., Graham, Y., ... & Zampieri, M. (2019). Findings of the 2019 Conference on Machine Translation (WMT19). In *Proceedings of the Fourth Conference on Machine Translation, Volume 2: Shared Task Papers*, Florence, Italy. Association for Computational Linguistics.

Bisong E. (2019) Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berkeley, CA: Apress.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the Role of BLEU in Machine Translation Research. In *11th Conference of the European Chapter of the Association for Computational Linguistics*, 249-256.

Cettolo, M., Girardi, C., and Federico, M. (2012). WIT3: Web Inventory of Transcribed and Translated Talks. In *Proceedings of EAMT 2012*, Trento, Italy.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv preprint arXiv:1409.1259.*

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv preprint arXiv:1412.3555.*

Darwin, C. M., & Gray, L. S. (1999). Going After the Phrasal Verb: An Alternative Approach to Classification. *TESOL Quarterly, 33*(1), 65-83.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805.*

Domingo, M., Garcıa-Martınez, M., Helle, A., Casacuberta, F., & Herranz, M. (2018). How Much Does Tokenization Affect Neural Machine Translation? *arXiv preprint arXiv:1812.08621.*

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14*(2), 179-211.

Gage, P. (1994). A New Algorithm for Data Compression. *C Users Journal, 12*(2), 23-38.

Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. *Synthesis lectures on human language technologies, 10*(1), 1-309.

Guillou, L., Hardmeier, C., Smith, A., Tiedemann, J., & Webber, B. (2014). ParCor 1.0: A Parallel Pronoun-Coreference Corpus to Support Statistical MT. In *Proceedings of LREC 2014,* Reykjavik, Iceland.

Harris, Z. S. (1945). Discontinuous Morphemes. *Language, 21*(3), 121-127.

Haspelmath, M. (2002). Understanding Morphology. London: Arnold.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735-1780.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-Strength Natural Language Processing in Python. *doi: 10.5281/zenodo.1212303.*

Jawahar, G., Sagot, B., & Seddah, D. (2019). What Does BERT Learn About the Structure of Language? In *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy.

Jurafsky, D., & Martin, J. (2020). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd Edition Draft. Retrieved in June, 2021, from https://web.stanford.edu/~jurafsky/slp3/.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., ... & Herbst, E. (2007). Moses: Open-Source Toolkit for Statistical Machine Translation. In *ACL Demo Session*.

Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical Phrase-Based Translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 127-133.

Kudo, T. (2018). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. *arXiv preprint arXiv:1804.10959.*

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, *19*(2), 313-330.

Mareček, D., & Rosa, R. (2019). From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions. *arXiv preprint arXiv:1906.01958.*

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, 51-56.

Ng, N., Yee, K., Baevski, A., Ott, M., Auli, M., & Edunov, S. (2019). Facebook FAIR's WMT19 News Translation Task Submission. *ArXiv, abs/1907.06616.*

Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., & Auli, M. (2019). fairseq: A Fast, Extensible Toolkit for Sequence Modeling. *arXiv preprint arXiv:1904.01038.*

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311-318.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-training (OpenAI Pre-print).

Schuster, M., & Nakajima, K. (2012, March). Japanese and Korean Voice Search. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5149-5152.

Sennrich, R., Haddow, B., & Birch, A. (2015). Neural Machine Translation of Rare Words with Subword Units. *arXiv preprint arXiv:1508.07909.*

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *arXiv preprint arXiv:1409.3215.*

Thim, S. (2012). *Phrasal verbs: The English Verb-Particle Construction and its History*. Walter de Gruyter.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All You Need. *arXiv preprint arXiv:1706.03762.*

Vestergard, T. (2019). *Prepositional Phrases and Prepositional Verbs: A Study in Grammatical Function.* Walter de Gruyter GmbH & Co KG.

Vig, J., & Belinkov, Y. (2019). Analyzing the Structure of Attention in a Transformer Language Model. *arXiv preprint arXiv:1906.04284.*

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … & van Mulbregt, P. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods, 17*(3), 261-272.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2019). HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv preprint arXiv:1910.03771.*

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google’s Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation. *arXiv preprint arXiv:1609.08144.*

ONLINE TOOLS AND RESOURCES

Cambridge Online Dictionary (n.d.). Accessed in June, 2020, at https://dictionary.cambridge.org/.

Collins Online Dictionary (n.d.). Accessed in June, 2020, at https://www.collinsdictionary.com/.

Google Translate (n.d.). Accessed in June, 2020, at https://translate.google.com/

Google Books Ngram Viewer (n.d.). Accessed in June, 2020, at https://books.google.com/ngrams.

Langenscheidt Online Dictionary (n.d.). Accessed in June, 2020, at https://en.langenscheidt.com/.

Merriam-Webster.com Dictionary (n.d.). Accessed in June, 2020, at https://www.merriam-webster.com/.

APPENDICES

Appendix A. TED Talks Documents Available at ParCor Collection

Title	Tokens		Parallel
	English	German	Sentences
Bill Gates on Energy: Innovating to Zero!	5,371	4,775	259
Aimee Mullins: The Opportunity of Adversity	3,414	3,430	143
Daniel Kahneman: The Riddle of Experience vs. Memory	3,564	3,566	181
Gary Flake: Is Pivot a Turning Point for Web Exploration?	1,280	1,163	65
James Cameron: Before Avatar ... a Curious Boy	3,265	3,054	172
Dan Barber: How I Fell in Love With a Fish	2,988	2,921	214
Eric Mead: The Magic of the Placebo	1,788	1,768	112
Jane McGonigal: Gaming Can Make a Better World	4,354	3,947	251
Robert Gupta: Music is Medicine, Music is Sanity	1,002	989	43
Michael Specter: The Danger of Science Denial	3,644	3,531	255
Tom Wujec: Build a Tower, Build a Team	1,301	1,161	81

Note. The table is adapted from Guillou et al. (2014, p.2).

Appendix B. Dependency Trees

Figure B1

Dependency Tree for the Sentence "John is going on a trip." (parsed with the spaCy library, visualized with the displacy tool)

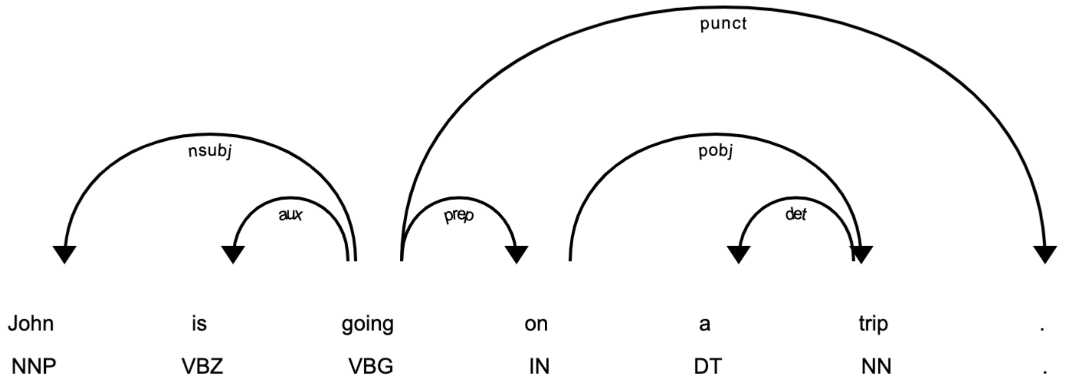
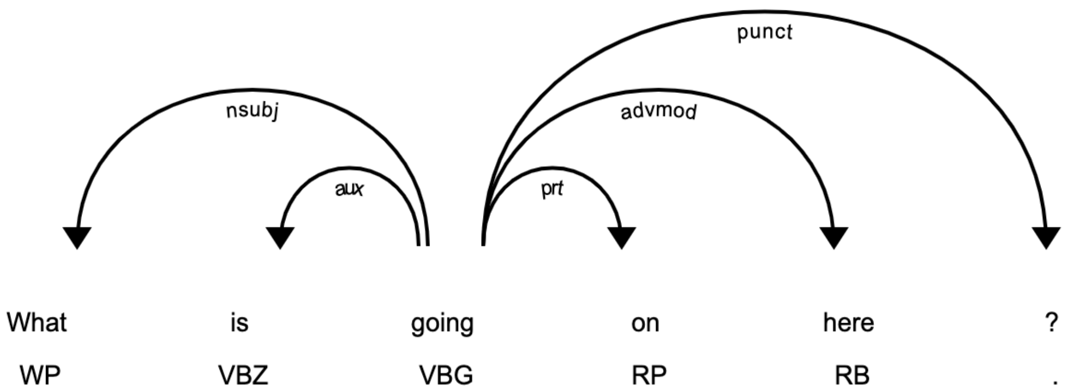


Figure B2

Dependency Tree for the Sentence "What is going on here?" (parsed with the spaCy library, visualized with the displacy tool)



Appendix C. Contingency Tables

Table C1

Contingency table for the variables "Input Type (German)" and "Output Type (English)"

Output Type (English)	Phrasal verb	Non-phrasal verb
Input Type (German)		
Separable verb	38	19
Inseparable verb	4	12
Verb with no prefix	9	25

Table C2

Combined contingency table for the variables "Input Type (DE)" and "Output Type (EN)"

	EN Phrasal Verb	EN Non-Phrasal Verb
DE Separable Verb	38	19
DE Non-separable Verb	13	37

Table C3

Contingency table for the variables "Input: Is Separated?" and "Output Type (English)"

Output Type (English)	Non-phrasal verb	Phrasal verb
Input: Is Separated?		
No	10	23
Yes	9	15

Appendix D. Translation Frequencies

Figure D1

Translations of "emittieren" from German to English and Their Frequencies in Public Documents by Google Translate

Translations of emittieren

Verb		Frequency
emit	emittieren, aussenden, abgeben, ausstrahlen, abstrahlen, ausstoßen	
issue	erteilen, ausstellen, ausgeben, erlassen, emittieren, herausgeben	
launch	starten, lancieren, einführen, gründen, auf den Markt bringen, emittieren	

Figure D2

Translations of "steigen" from German to English and Their Frequencies in Public Documents by Google Translate

Translations of steigen

Verb		Frequency
rise	steigen, ansteigen, aufsteigen, aufstehen, aufgehen, sich erheben	
increase	erhöhen, steigern, vergrößern, steigen, zunehmen, verstärken	
climb	klettern, steigen, besteigen, ersteigen, erklettern, aufsteigen	
go up	steigen, hinaufgehen, ansteigen, aufsteigen, hinauffahren, aufgehen	
ascend	aufsteigen, steigen, besteigen, hinaufsteigen, auffahren, ansteigen	
soar	steigen, schweben, aufsteigen, hochschnellen, aufstreben, hochragen	
step	treten, gehen, steigen, tanzen	
mount	montieren, anbringen, aufstellen, besteigen, steigen, aufsteigen	
move up	rücken, aufsteigen, steigen, aufrücken, vorrücken, auffahren	
lift	heben, anheben, erheben, aufheben, hochheben, steigen	

Appendix E. Google Ngrams

Figure E1

Comparison Chart of the Frequencies of "get to the goal" and "go to the goal" via Google Books Ngram Viewer

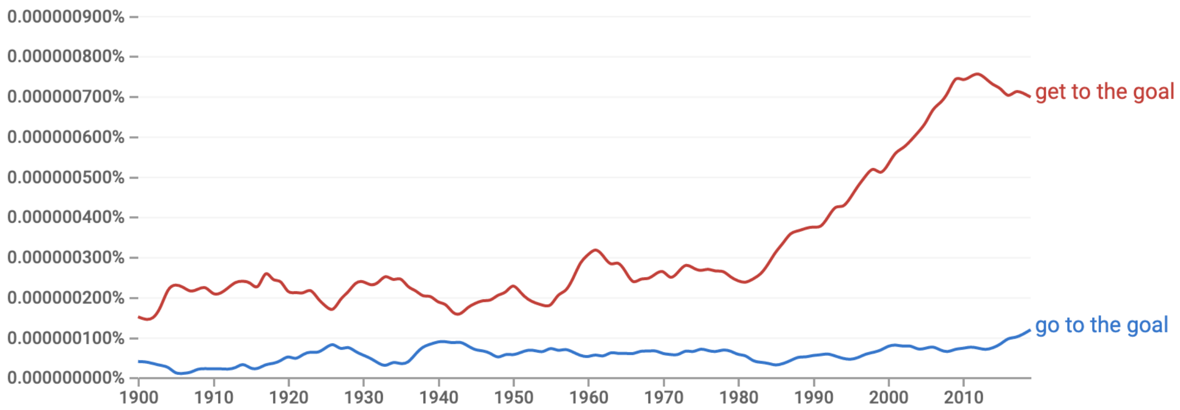


Figure E2

Comparison Chart of the Frequencies of "get to the goal" and "go to the goal" via Google Books Ngram Viewer